

# IF704 - Processamento de Linguagem Natural

Erick de Almeida Oliveira Riso

25 de outubro de 2018

## 1 Introdução

O Processamento de Linguagem Natural (do inglês Natural Language Processing, às vezes também referida como PLN) é uma disciplina eletiva do curso de graduação de Ciência da Computação da Universidade Federal de Pernambuco (UFPE), ministrada pelo professor Hansenclever Bassani<sup>1</sup>. É considerada uma subárea da Inteligência Artificial, e tem como objetivo tornar possível a interpretação das línguas humanas naturais pelos sistemas computadorizados. O Processamento de Linguagem Natural incorpora várias técnicas oriundas de disciplinas diferentes, que vão desde a Linguística, através de itens como a semântica, sintaxe e lexicologia, até a Estatística e métodos de aprendizagem de máquina.

## 2 Relevância

A comunicação humana geralmente envolve fatores muito complexos, como contexto e especificidades regionais, e está sujeita à ambiguidades. Os computadores, todavia, geralmente trabalham com dados estruturados e linguagem precisa. Conciliar esse problema é um dos desafios do Processamento de Linguagem Natural. Grandes empresas têm dado muita importância ao ramo, já que esta área pode auxiliar a resolução de muitos problemas hodiernos da computação, como a análise de Big Data e o aperfeiçoamento de sistemas de aprendizado de máquina. Muitos analistas apostam, inclusive, que o Processamento de Linguagem Natural seja parte integrante do que virá a ser “a grande novidade” da tecnologia da informação.<sup>2</sup>

Hoje, muitas aplicações utilizam processamento de linguagem natural, e entre elas podemos destacar:

- Tradução automática de textos
- Reconhecimento de voz
- Filtragem de SPAM
- Reconhecimento ótico de caracteres
- Análise de sentimentos
- Sumarização de documentos
- Correção ortográfica
- Chatbots e assistentes pessoais

---

<sup>1</sup>BASSANI, Hansenclever. *Página oficial da disciplina.* Disponível em: <<http://www.cin.ufpe.br/hfb/pln/index.html>>. Acesso em: 23 de outubro de 2018

<sup>2</sup>DEANGELIS, Stephen F. *The Growing Importance of Natural Language Processing.* Disponível em: <<https://www.wired.com/insights/2014/02/growing-importance-natural-language-processing/>> (em inglês). Acesso em: 25 de outubro de 2018

## 2.1 Pontos positivos e negativos

Entre os pontos positivos da disciplina, evidencia-se:

- Área desfruta de relevância para grandes empresas
- Grande interdisciplinaridade com outras áreas, como Inteligência Artificial, Aprendizado de Máquina, Estatística, Linguística e até mesmo Psicologia
- Liberdade para o estudante poder testar e explorar, já que não há apenas uma única forma de se resolver um determinado problema

Já entre os pontos negativos, é possível salientar:

- Pouco material de estudo disponibilizado em língua portuguesa
- Se comparado à outras línguas, como o inglês, o português ainda não tem tanto destaque enquanto objeto de estudo da disciplina, o que se reflete em menos ferramentas para processamento e análise de tal língua.

## 3 Relação com outras disciplinas

Disciplinas	Relação
Processamento de Voz (IF759)	Um dos ramos do Processamento de Linguagem Natural é o Processamento de Voz. É esperável, portanto, que esta disciplina tenha o IF704 como pré-requisito.
Sistemas Inteligentes (IF684)	Uma importante demanda na criação de sistemas inteligentes é a de interagir e tentar compreender o interesse dos usuários. Quando esta tarefa depende da manipulação de linguagem humana, o Processamento de Linguagem Natural pode ser um aliado.
Introdução à Programação (IF669)	Visto que o Processamento de Linguagem Natural lida com a criação de programas e sistemas para solucionar problemas de interpretação de linguagem humana, habilidades em programação são imprescindíveis.
Estatística e Probabilidade para Computação (ET586)	A Inteligência Artificial se vale de muitos conceitos da estatística, como a probabilidade, para a tomada de decisões. No caso específico do Processamento de Linguagem Natural, ela pode ser utilizada para definir o contexto de uma expressão em um texto específico, por exemplo.
Álgebra Vetorial e Linear (MA531)	Muitas noções da álgebra linear e vetorial são utilizadas em modelos para processamento de linguagem natural. Por exemplo, o espaço semântico, ferramenta empregada para representar o significado das palavras, recorre-se aos conceitos de espaço vetorial e vetor.

## Referências

- [1] Prabhakar Raghavan Christopher D. Manning and Hinrich Schutze. *An Introduction to Information Retrieval*. Cambridge University Press, 2009.
- [2] Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. Prentice Hall, 2008.
- [3] Ewan Klein Steven Bird and Edward Loper. *Natural Language Processing with Python*. O'Reilly Media, Inc, 2009.